When cognition turns vicious: Heuristics and biases in light of virtue epistemology

When Cognition Turns Vicious: Heuristics and Biases in Light of Virtue Epistemology

Peter L. Samuelson,

Fuller Theological Seminary School of Psychology

Ian M. Church

Saint Louis University

Abstract

In this paper we explore the literature on cognitive heuristics and biases in light of virtue epistemology, specifically highlighting the two major positions—agent-reliabilism and agent-responsibilism (or neo-Aristotelianism)—as they apply to dual systems theories of cognition and the role of motivation in biases. We investigate under which conditions heuristics and biases might be characterized as vicious and conclude that a certain kind of intellectual arrogance can be attributed to an inappropriate reliance on Type 1, or the improper function of Type 2 cognitive processes. By the same token, the proper intervention of Type 2 processes results in the virtuous functioning of our cognitive systems (agent-reliabilism). Moreover, the role of motivation in attenuating cognitive biases and the cultivation of certain epistemic habits (a search for accuracy, being accountable for one's judgments, the use of rules of analysis, and exposure to differing perspectives) points to the tenets of agent-responsibilism in epistemic virtue. We identify the proper use of Type 2 cognitive processes and the habits of mind that attenuate biases as demonstrations of the virtue of intellectual humility. We briefly explore the nature of these habits and the contribution of personality traits, situational pressures, and training in their cultivation.

§1. Introduction

Any account of knowing has, as its starting point, the capacity and desire to know the "truth," through whatever normative criteria the "truth" can be established. Yet, human beings are notoriously (and apparently naturally) disposed to over-estimate their capacity to know the truth and under-estimate their weaknesses (Chaiken, Wood, & Eagly, 1996; Dunning, Leuenberger, & Sherman, 1995; Evans, 2007; Gilovich, Griffin, & Kahneman, 2002; Kunda, 1990; Pronin, Berger, & Molouki, 2007; Stanovich & West, 1997; Wegener & Petty, 1997). Indeed, the evidence is clear that there is a strong tendency even to under-estimate our liability to such biases (Pronin & Kugler, 2007)! Furthermore, we are susceptible to all sorts of biases that make knowing difficult. For example, we tend to favor evidence or data received early in our inquiries (Kruglanski, Dechesne, Orehek, & Pierro, 2009) and we tend to discount the weight of evidence that counts against hypotheses we endorse (Nickerson, 1998). Second, evolutionary psychologists have offered some intriguing arguments that these dispositions are embedded within our cognitive architecture in ways that can systematically lead us to biased thinking, in some cases for adaptive reasons (Mercier & Sperber, 2011). Our intention here is to apply some principles and distinctions from virtue epistemology to the literature on cognitive heuristics and biases to gain a better understanding of the promise and perils of our cognitive systems. We will examine questions of how and under what circumstances these heuristics and biases can be characterized as epistemic vices and how the specific epistemic virtue of intellectual humility may help mitigate these vices and steer us toward the fundamental epistemic goal of "truth-tracking.[1]

§2. Virtue Epistemology

Distinctively, virtue epistemology places intellectual virtue at the heart of knowledge.[2] Virtue epistemologists predominantly focus on the agent himself or herself: on whether he or she has the right sort of epistemic character, the right sort of cognitive faculties, whether he or she is epistemically virtuous or not. To put it roughly, instead of focusing on the beliefs of agents (whether or not they are justified, rational, etc.), virtue epistemologists predominantly focus on the agent himself or herself: on whether he or she has the right sort of epistemic character, the right sort of cognitive faculties, whether he or she is epistemically virtuous or not. To be sure, other theories of knowledge will give some account of epistemic virtues, norms, and values—good memory, intellectual courage, etc.—but usually ancillary to other epistemic terms or concepts. The radical claim that virtue-theoretic accounts of knowledge make, however, is that epistemic virtues, norms, and values should be the primary focus of epistemology (see Pritchard (2005), p. 186; Greco (2010), pp. 17-46).

Virtue epistemology, so defined, has developed by and large into two distinct schools: agent-reliabilism and agent-responsibilism or neo-Aristotelianism. The primary difference between the schools is their application of "virtue" terminology. Agent-reliabilism, being modeled along reliabilist lines, applies virtue terminology in regard to faculties, in the same way we might talk about a virtuous knife. In other words, just as we might call a knife virtuous if it does what it is supposed to do (cut things, be sharp, etc.), agent-reliabilism calls various cognitive faculties such as memory, perception, etc.,

virtuous insofar as they are reliably functioning the way they are supposed to. That is, agent-reliabilism focuses on the reliable functioning (virtuous functioning) of a given agent's cognitive faculties. Neo-Aristotelianism, on the other hand, applies virtue terminology in a way with which we are perhaps more familiar: in terms of specific character traits such as open-mindedness, intellectual courage, intellectual perseverance, etc.

§2.1. Agent-reliabilism.

Agent-reliabilism virtue epistemologies developed out of a dissatisfaction with what is called process-reliabilism: the view that, to put it roughly, S knows a true proposition *p* if and only if *p* was formed by a reliable process. There are some serious concerns for such a view (e.g., the "generality problem");[3] however, the agent-reliabilists' primary concern was that knowledge ascriptions based on reliable processes do not always appropriately involve a given agent: that process-reliabilism seems to allow agents to "possess knowledge even though the reliability in question in no way reflects a cognitive achievement on their part" (Pritchard, 2005, p. 187).

Another way that knowledge ascriptions based on reliable processes do not always appropriately involve a given agent is in the case of reliable cognitive malfunctions. Consider a case originally developed by Alvin Plantinga in which our protagonist has a brain lesion that causes him to believe he has a brain lesion:

> **Brain Lesion:** Suppose . . . that S suffers from this sort of disorder and accordingly believes that he suffers from a brain lesion. Add that he has no evidence at all for this belief: no symptoms of which he is aware, no testimony on the part of physicians or other expert witnesses, nothing.

> (Add, if you like, that he has much evidence against it, but then add also that the malfunction induced by the lesion makes it impossible for him to take appropriate account of this evidence.) Then the relevant [process] will certainly be reliable but the resulting belief—that he has a brain lesion—will have little by way of warrant for S. (Plantinga, 1993, p. 199)[4]

Again, though S's belief that he has a brain lesion is formed via a reliable process we would not ascribe knowledge to it since it is only formed out of a glitch in S's cognitive equipment. It is simply accidental that the brain lesion causes S to form the said belief, and as such S had nothing to do with its formation. Though there may be some worries as to what constitutes a given person's cognitive equipment (Why is S's brain lesion not a part of his cognitive equipment? What if he had the lesion since birth?), we nevertheless have strong intuitions that beliefs formed as a direct result of a cognitive malfunction cannot be knowledge.[5] S's belief that he has a brain lesion cannot be knowledge simply because, though formed via reliable process, the agent, S, was not appropriately involved in its formation.

§2.2. Neo-Aristotelianism (Responsibilism).

Neo-Aristotelianism virtue epistemologies, in contrast, are not modeled after reliabilism. Instead of focusing on whether or not a given agent's epistemic faculties are functioning properly and reliable, neo-Aristotelian virtue epistemologists tend to focus more on the agent's epistemic character and epistemic responsibilities. Consider the following case:

> **Chicken Sexer:** Naïve and Reflective are both chicken sexers. Their job is to look at baby chickens, determine their genders, and then segregate the

> chickens accordingly, putting male chicks in one box and female chicks in another. Both Naïve and Reflective are equally good at their job; both are highly reliable at determining the gender of the baby chickens. There is, however, one important difference between them: Naïve has no idea how he is able to correctly determine the gender of the chicks; he "just does it." Reflective, on the other hand, is very much aware of how he makes such a judgment, by looking for a certain pattern in the chick's feathers. Does Naïve's ignorance affect his ability to know "that's a male chick," "that's a female chick," etc.?

Whereas agent-reliabilism virtue epistemologists would generally deny that Naïve's ignorance affects his ability to know (after all, his cognitive equipment is highly reliable), neo-Aristotelianism virtue epistemologists would cry foul (or fowl). In focusing on epistemic character traits instead of epistemic faculties, neo-Aristotelianism tends to "stress that agents should not only exhibit reliable cognitive traits but that they should also be in a position to take… reflective responsibility for their true beliefs" (Pritchard, 2005, pp. 194-195). Naïve may very well have reliable cognitive faculties, but, for the neo-Aristotelianism virtue epistemologist that is not enough; in order to be properly said to *know* the gender of the chicks, Naïve needs to exhibit more conscientiousness and take more epistemic responsibility.

This focus on epistemic responsibility ties into another common hallmark of neo-Aristotelianism, namely, a closer correspondence between epistemology and moral philosophy. As we noted earlier, the primary objects of interest for the neo-Aristotelian virtue epistemologist are not cognitive faculties (as it was with the agent-reliablist) but

rather intellectual character traits (intellectual courageousness, open-mindedness, etc.), whether or not a given agent is of the right sort of epistemic character. And being of the right sort of epistemic character often means (at the very least) not only reliably reaching virtuous ends/*teloi* but also being virtuously motivated. In other words, in order to be of the right sort of epistemic character, not only do you need to be the sort of person who regularly hits upon the truth, but you need to hit upon the truth for the right reasons (e.g. because you were intellectually courageous as opposed to simply lucky). Not only is neo-Aristotelianism interested in your veritic reliability, neo-Aristotelianism is interested in what sort of person— what sort of epistemic character—you should be. And all of this gives epistemology a distinctive moral dimension. Indeed, some neo-Aristotelian virtue epistemologists (e.g. Zagzebski, 1996) have even gone so far as to elucidate epistemic virtues as a subset of moral virtues.

The primary difference between agent-reliabilism virtue epistemology and neo-Aristotelian virtue epistemology that we want to focus on is their divergent accounts of intellectual virtue: agent-reliablists roughly explicating intellectual virtue in terms of cognitive faculties or cognitive competencies (faculty virtues) and neo-Aristotelians roughly explicating intellectual virtues in terms of character traits and motivation (character-virtues). Most often, neo-Aristotelianism can be seen as requiring something more than the agent-reliabilist: requiring not only that a given agent have reliable cognitive faculties (like Naïve in Chicken Sexer), but that they also be of the right sort of epistemic character. However, this isn't necessary the case; neo-Aristotelianism (as we are currently conceiving of it) could be seen as simply requiring something else, not something more. For example, a neo-Aristotelian could conceivably understand an

intellectually virtue—character virtues such as intellectual courage, intellectual steadfastness, etc.—in such a way that it did not necessary require reliability; such a neo-Aristotelian might ascribe knowledge to an agent who manifests such a virtue even when the faculties or competencies at play are unreliable.

§3. The Virtuous Knower

The distinction in virtue epistemology between agent-reliabilism and agent-responsibilism (or neo-Aristotelianism) provides a useful framework for examining the psychological literature on heuristics and biases. On the one hand, we see a concern for the reliability (or lack thereof) of our cognitive systems in the literature that deals with dual-process theories of cognition, heuristics, and biases (Epstein, Lipson, Holstein, & Huh, 1992; Evans, 2007; Gilovich, 1991; Gilovich, et al., 2002; Kahneman, 2011; Sloman, 2002; Stanovich, 1999). On the other hand, there is a concern for responsibilism reflected in a focus on the role of motivation in cognition and cognitive bias (Chaiken, et al., 1996; Dunning, et al., 1995; Kunda, 1990; Pronin, et al., 2007; Wegener & Petty, 1997). It is not as if those concerned with heuristics and biases ignore the role of motivation (Evans, 2007; Stanovich & West, 1997) nor do those primarily interested in the role of motivation ignore cognitive mechanisms and capacity (Chaiken, et al., 1996). It is rather a matter of emphasis. Nevertheless, these distinctions can serve to provide a better understanding of how to best reach the epistemic goal of "truth-tracking" in light of the heuristics and biases of our cognitive system.

§3.1. Reliabilism: Heuristics and Biases.

§3.1.1. *Dual process theories*.

From a cognitive science perspective, it could be said that if one really knew about the systematic heuristics and resulting biases that appear to be a part of our shared cognitive mechanisms, one could not help but be intellectually humble (or at least humbled). Research into these cognitive heuristics that speed processing and the resulting chronic and systematic biases is extensive, not only in the various fields of psychology, but in other disciplines as well (see Kahneman, 2011, for one of the latest summaries). Out of research into heuristics and biases has grown a number of what are called "dual-process" theories of human cognition.[6] While each theory has different names for and different categories assigned to each process, these theories broadly share a distinction between fast, automatic, and intuitive processes, called Type 1 (also known as System 1) processes, and slow, deliberative, and analytic processes known as Type 2 (also known as System 2) processes (Kahneman, 2011; Kahneman & Frederick, 2002; Stanovich, 1999; Evans & Stanovich, 2013)[7].

The central characteristic of Type 1 processing is its automaticity – processes that do not make much demand on working memory and are not governed by what Evans & Stanovich, (2013) call "controlled attention" (p. 236). Because of this, Type 1 processing has many correlated (but not necessarily defining) features, such as: rapid processing that sometimes does not reach consciousness, associative processing that reacts to stimuli with minimal cognitive load, and processing that relies on thoughts that are most readily available and dependent on personal experience. Many of the heuristics (mental shortcuts that speed processing) and resulting biases are therefore also correlates of Type 1 processing, though Type 2 processing can also result in systematic biases.[8]

Type 2 processing is characterized by hypothetical thinking: thinking that is "decoupled" from an individual's representation of reality. This "cognitive decoupling" is defined as "the ability to distinguish supposition from belief and to aid rational choices by running thought experiments" (Evans & Stanovich, 2013, p. 236). Stanovich posits two "modes" of thinking within Type 2 processing: "algorithmic" and "reflective." (Evans & Stanovich, 2013; Stanovich 2009). Although more effortful and requiring more working memory than Type 1 processing, the individual's capacity for engaging these modes of thinking that are constitutive of Type 2 processing exists on a continuum such that some may execute them with alacrity while others will take more time and deliberation. The algorithmic mode of processing suppresses the automatic responses of Type 1 processing, decoupling from the current representation of the world, in order to compare it to some kind (or kinds) of secondary representation(s). The source of these other representations could come from rule-based thinking, from simulations and hypotheticals based on experiences (or other knowledge sources), and/or from alternative models that are learned, among others. Because the current representation must be held in mind while being compared to other possible representations to find the best one (requiring effort and a load on working memory) algorithmic processing is highly correlated with fluid intelligence, a key factor in intelligence testing (Stanovich, 2009). The "reflective" mode of Type 2 processing is characterized by higher forms of cognitive regulation and reflects the goals and epistemic values of an individual. Labeled as "thinking dispositions" (Stanovich & West, 1997, p. 343) they correlate with such traits and tendencies as open-mindedness, thinking through problems and weighing

consequences before taking action, and gathering sufficient evidence before making conclusions, among others (Evans & Stanovich, 2013).

From a virtue reliablist point of view, epistemic vice cannot reside alone in either Type 1 or Type 2 processing, though most biases are attributed to an over-reliance on Type 1 processing (Evans, 2007; Kahneman & Frederick, 2002; Evans & Stanovich, 2013). Type 1 processing can be and is quite reliable and often hits on the truth. The epistemic vice is found in the breakdown of the relationship between the two types of processing and virtue is attained when each plays its appropriate function in the pursuit of epistemic goods such as truth, accurate representation of reality, and etc. The relationship has been described as "default/interventionist" (Kahneman, 2011; Evans & Stanovich, 2013), that is, Type 1 processing, because of its efficiency, is the default type of processing until and unless Type 2 thinking, which monitors all thought processes, determines deeper processing is needed and intervenes to decouple the Type 1 representation and hypothetically entertain other possible representations in order to find the epistemic good (truth, accuracy, knowledge, etc., Stanovich, 1999, 2009, Evans & Stanovich, 2013). If Type 2 processing is lazy, negligent, distracted, or overly focused on the self, biases will go unnoticed (Kahneman, 2011). Indeed, there is broad consensus that biases are effectively reduced through effortful, deliberate, analytic, Type 2 processes (Evans, 2007; Kahneman, 2011; Sloman, 2002; Stanovich & West, 1997; Wilson, Centerbar, & Brekke, 2002). That the automatic Type 1 processing does not hit upon the truth is not epistemically vicious in and of itself, but would be if Type 2 processing, for whatever reason, fails to correct or amend the Type 1 representation when

it is distorted. Both need to fail in their proper function for an epistemically vicious result.[9]

Stanovich (2009) has built a basic taxonomy that lays out thinking errors and biases along virtue reliablist lines. First there are the "cognitive miser" errors (p. 74-75) which means, as the name implies, the cognitive system (specifically Type 2 thinking) did not expend enough resources by 1) failing to de-couple and therefore going with Type 1 processing even though biased, 2) decoupling but failing to override Type 1 processing when it should have been or 3) decoupling but perseverating on only one alternative representation when more comparisons are needed. In each case Type 2 thinking did not perform its proper function in the pursuit of epistemic goods. Then there are "mindware" problems, which can be manifested as a "mindware gap," (p. 75) that is, a gap in the learning necessary for Type 2 processing to recognize the need for decoupling and further reflection. This can include gaps in the knowledge of rules, of procedures and strategies for thinking, of probability, of specific domain knowledge, among others. Then there is "mindware contamination (p. 76)" which includes learned rewards and punishments for not engaging in decoupled thinking, using representations of the world from a egocentric perspective in decoupling, and using knowledge structures that are wrong or misguided in decoupling. As the labels cognitive miser" and "mindware" imply, the cognitive system is not working virtuously and is prone to epistemic vice due to improper functioning such as failing to engage Type 2 processing when necessary, gaps in knowledge, and contamination of mental representations used in Type 2 processing.

The heuristics of Type 1 thinking evaluate ideas from within one's own perspective and therefore favor ideas that are readily accessible, easily discerned, and conforms to

prior experience (Dunning, Meyerowitz, & Holzberg, 2002; Sanna, Schwarz, & Stocker, 2002; Pronin et al., 2007; Stanovich & West, 2007). Type 2 processing can also favor what one already believes, knows, and intuits even as it decouples from Type 1 representations (Stanovich , 2009, calls this "egocentric processing," p. 78). While in many instances this self-reliant thinking is adequate, problems and biases arise when reliance on what one knows and intuits does not provide enough information (or the right kind of information) for the task, which, in turn, produces biased judgments. Many of these biases are well known and well documented. There are (among others):

- The confirmation bias: The tendency to seek confirmation for opinions and beliefs already held and to ignore disconfirming evidence (Nickerson, 1998). Even academic psychologists are not immune. They rate studies with findings consistent with their prior beliefs more favorably than studies with conclusions inconsistent with their beliefs (Hergovich, Schott, & Burger, 2010).
- The hindsight bias: People's predictions of events are remembered as more accurate after the fact than they really were. In predicting the outcome of the German Bundestag elections of 1998, subjects remembered—after the election—their predictions 4 months earlier as 25% closer to the actual results than they had originally predicted (Blank, Fischer, & Erdfelder, 2003).
- The anchoring and adjustment heuristic: Cognitive anchors affect people's judgments under conditions of uncertainty. When asked to estimate when George Washington was elected president, participants began with a known date (the Declaration of Independence, 1776) and adjusted from there to the unknown date (Epley & Gilovich, 2002).

- The my-side bias: People are biased toward their own opinions and point of view when evaluating evidence, generating evidence and testing hypotheses, regardless of their level of intelligence. They also have trouble assessing conclusions that conflict with their knowledge of the world, However, when explicitly instructed to consider other points of view, people high in general intelligence and possessing certain thinking dispositions (e.g. actively open-minded thinking) exhibited less biased thinking (Stanovich & West, 2007).

§3.2. Responsibilism: Motivation, Goals, and Values in Cognition.

As noted above, responsibilism (neo-Aristotelian virtue epistemology) is not as interested in the reliability of our cognitive mechanism (compared to the reliabilists) as they are in whether or not a given agent has of the right sort of epistemic character. A person with the right sort of epistemic character will not only reliably reach virtuous ends but must also be virtuously motivated. Psychologists, too, are interested in epistemic motivation and the goals and values of cognition. Many of them use dual system theories of cognition as their framework. There is some disagreement, however, about whether these are character traits in the sense that they are an enduring part of one's personality, or whether these traits are largely situational, i.e. dependent on the agent's state.

In light of a cognitive system that is often times is prone to biases because of a failure to of Type 2 processing to decouple from Type 1 representation when necessary, one avenue of exploration into virtuous knowing would be to investigate processes and actions that attenuate biases. From a virtue-responsibilist point of view, the vice of cognitive bias lies in this inability (or perhaps refusal) to decouple from the singular representation of reality that Type 1 processing provides. In this way, the cognitive

system, in its insistence to stay with Type 1 representations and to not make algorithmic or hypothetical comparisons to other possible representations, could be characterized as "self-centered." Therefore, reducing or eliminating these biases might involve some kind of engagement with an "other;" someone or something that "de-centers" the cognitive system, engages decoupling, and entertains different ways of thinking and other points of view. Indeed a review of the literature reveals several types of "other-centered" thinking as effective techniques for reducing biases: the use of rules of analysis (a process used by many others to arrive at a more consensual judgment), a search for accuracy (representing reality that is shared by others), a need to be accountable for one's judgments (to defend one's thoughts to another), and exposure to differing perspectives (seeing things from another's point of view). The presence or absence of these factors often hinges on motivation and the epistemic goals and values of thinking agents and thus favors the responsibilist or neo-Aristotelian account of virtue epistemology.

§3.2.1. ***Rule-based thinking.***

Directing one's attention to processes, objective criteria, and rules of analysis can aid in reducing systematic bias. For example, when personal traits are ambiguous (e.g., whether or not one is intelligent or a good driver), people draw on idiosyncratic definitions of traits and abilities to assess themselves as better than average at a certain task or trait. Once given criteria of judgment, however, they are more accurate in their assessment relative to their peers (Dunning, et al., 2002). Evans (2007) demonstrates that when questions are asked in a way that engages analytic reasoning processes, biases are reduced. For example, his research has shown that the usual matching bias evident in the original form of the Wason card selection task is eliminated when the rule is highlighted

over the surface features, making the need to falsify in order to rightly complete the task explicit. Stanovich (1999) has noted that many of the tasks involved in heuristics and biases depend on an understanding and mastery of normative rules of thought (logic and statistics) which is the purview of the algorithmic function of Type 2 processing. Those attuned to rule-based, algorithmic thinking perform better on these tasks. Understanding can come through training (as the research reported in Stanovich (1999) demonstrates), through the framing and presentation of the problem (Evans, 2007), through cognitive ability (high analytic reasoning capability) and through the possession of certain open and flexible thinking dispositions that "serve the ends of epistemic rationality" (Stanovich & West, 1998, p. 180).

§3.2.2. ***Accuracy and accountability.***

Situations that call for accuracy in judgment promote slower, more deliberative thinking. Kunda (1990), in an analysis of the work on accuracy-driven reasoning, concluded that when "people are motivated to be accurate, they expend more cognitive effort on issue-related reasoning, attend to relevant information more carefully, and process it more deeply, often using more complex rules" (p. 481). For example, when people know they will be judged on accuracy, they are more accurate in evaluating their own abilities (Armor & Taylor, 2002). However, Petty, Wegener, & White (1998) caution that motivation for accuracy may not be enough to attenuate bias, but that bias can also occur in the correction process. Nevertheless, Kruglanski & Mayseless (1987) report that a high need for accuracy (what they call a heightened fear of invalidity) motivates people to seek comparison with those who disagree with them. The focus on accuracy, then, invites the thinking agent into a more "humble" epistemic posture

because (a) that agent may realize that he or she has a less than complete understanding and needs to seek more information, (b) it helps the agent to focus on what others might think of the same phenomenon, and (c) it will focus the agent on objective criteria about the phenomenon. The focus on accuracy and accountability emphasizes the situational factors that attenuate a generalized human tendency toward biased thinking.

Having to defend one's thoughts and judgments to others might include the need to be accurate, but also injects the notion of accountability, which promotes a more careful analysis of one's thoughts and arguments. Mercier & Sperber (2011), in a thorough examination of the many abstract reasoning tasks that are used to measure heuristics and biases, demonstrate that when the same reasoning tasks are set in the context of making an argument or defending a position, the reasoning of the participants is less biased and more complete. Moreover, they make the case that the confirmation bias, in the context of producing arguments to convince others of the rightness of one's beliefs, can produce a "cognitive division of labor." Because participants in a disagreement want to bring the best evidence they have found to support their beliefs, the confirmation bias serves them well collectively: together two disagreeing parties will tend to marshal relevant evidence in favor of their favoring positions. However, in the context of a discussion with others over the best evidence, or the best argument (the *process of evaluating* an argument, not its production), the confirmation bias no longer serves them well. Therefore, this "cognitive division of labor" works in the context of disagreement provided (and this is an important caveat) people have "a common interest in the truth" and go through the hard work of evaluating and selecting the best argument with the best evidence. Most people are quite good at the task of evaluating the best argument, both at the individual

and group level, when motivated to do so. When held accountable under a time pressure, however, people exhibit the primacy effect, giving more weight to information they received first (Kruglanski, et al., 2009).

The key to mitigating bias may be in taking the position of "a common interest in the truth." A study by Stanovich & West (1997) on the capacity of the individual to evaluate an argument fairly in light of previously held beliefs is informative in this regard. 349 college students completed a measurement called the Argument Evaluation Test (AET) in which they first indicated the strength of their beliefs regarding certain social and political issues and then (later in the testing process) evaluated the quality of the arguments of a fictitious individual on these same issues. Their analysis resulted in two groups, one with a high reliance on argument quality and a low reliance on prior belief and another with a low reliance on argument quality and a high reliance on prior belief. They compared the mean scores of the two groups on various measures including a composite score that measured "Actively Open-minded Thinking" (AOT) that indicated openness to belief change and cognitive flexibility. Those who showed a high reliance on argument quality by relying less on prior beliefs scored significantly higher on the AOT composite scale. This open and flexible thinking disposition held even when they controlled for cognitive ability.[10] Stanovich & West (1997) assert that these thinking dispositions can provide information about an individual's epistemic goals and values. For example, a disposition such as the willingness to change beliefs reflects the goal of getting as close to the truth as possible, or the disposition to carefully evaluate arguments indicates a value for accuracy. It might be fair to infer that those with a high reliance on argument quality had the epistemic goal of "a common interest in the truth." Thinking

dispositions function more like enduring traits that are less influenced by varying circumstances, though they are seen as "malleable" and therefore teachable skills (Baron, 1994).

§3.2.3. *Perspective taking.*

One aspect of an open and flexible thinking disposition that helps mitigate biases is the capacity to weigh evidence for and against a strongly held belief, including the opinions and beliefs of others who hold a position different from one's own. This requires a certain capacity for perspective taking: a movement from the focus on one's own thoughts to include the perceptions, thoughts and ideas of others. This adjustment can be as simple as considering alternative points of view and as complex as trying to assess another person's thoughts. Studies have found that asking people to consider the possibility that competing hypotheses are true is sufficient to undo the bias of one-sided thinking (Sedikides, Horton, & Gregg, 2007; Wilson, et al., 2002). The issue can be complicated, however, by the primacy effect. Jonas, Schulz-Hardt, Frey, & Thelen (2001) demonstrated that the order of the presentation of other points of view can impact de-biasing. The confirmation bias was stronger in a situation in which subjects were exposed to other points of view sequentially and lessened when subjects had access to all the points of view simultaneously. The sequential presentation encouraged the subjects to remain focused on their prior commitments, whereas those who had the information presented simultaneously were able to compare their beliefs to many points of view. This might lead to the conclusion that entertaining more alternatives would help attenuate bias, but Sanna, et al. (2002) have shown that in the case of hindsight bias more is not better. Subjects who were asked to produce 12 reasons why the British-Gurka war could have

come out differently were more susceptible to hindsight bias than those who produced only two reasons. They surmise that the difficultly of coming up with 12 counterfactual reasons makes alternative outcomes seem less likely and the actual outcome more so and puts the availability heuristic into play. However, those who came up with only two reasons performed as well as the control group who did not know the outcome of the war. Thus, in moderation, considering a counterfactual perspective helped attenuate hindsight bias.

Often bias is the result of a lack of perspective taking. Birch & Bernstein (2007) propose that a similar inability to take the perspective of naïve others is at the core of both children who are unable to successfully complete Theory of Mind (ToM) tasks and adults who exhibit hindsight bias. In each case the "knower" and naïve other do not share the same information because the knower is biased by the primacy of current knowledge and belief. On the other hand, perspective taking itself also is prone to bias. When assessing the behavior of others, people exhibit what is known as the correspondence bias, which is the tendency to make dispositional inferences from behaviors that can be mostly explained by the situations in which they are found. For example, when observing a person at a party that does not talk very much, you assume she is shy. Later you find out she is from another country and not very confident in her English speaking ability. A few months later you meet her again at a party and she is very talkative, even extroverted. It was being in foreign country without speaking the language that made her look shy at first, when, in fact, she has a disposition toward extroversion. This comes about, in part, because of what Gilbert & Malone (1995) call an "egocentric assumption" which includes an inability to "put (oneself) in someone else's epistemic shoes" (p. 26, i.e.

construe the situation as the actor does). However, when specifically prompted to see a situation as another sees it, egocentric biases can be attenuated. For example, Todd, Bodenhausen, Richeson, & Galinsky (2011) found that subjects who performed a perspective taking task showed less implicit or automatic racial bias (as measured by the personalized evaluative race IAT) than a control group that did not engage in perspective taking. One possible explanation the researchers offered was that since the self is the anchor in evaluating others, lessening the distance between the self and the person of another race through perspective taking allowed for more positive evaluations of the person of another race.

§4. Heuristics and Biases as Intellectual Arrogance.

In a limited sense we might say that heuristics and biases exhibit the vice of "intellectual arrogance" (IA) because, they result, in part, from an inability to 'decouple' from Type 1 representations and the thinker remains unable to leave his or her own perspective. There are also instances when decoupling occurs and Type 2 thinking is engaged but it remains focused on the self, resulting in egocentric thinking (Stanovich, 2009). In both cases, thought remains "self-centered." We see this exhibited in many of the biases identified in the literature. The self- centeredness is found in the general human tendency to use the self as an anchor against which the other is compared and the world is known (Dunning, Krueger, & Alicke, 2005; Guenther & Alicke, 2010). We are biased toward that which comes fastest and most easily to mind, (availability or representative heuristic, Kahneman & Frederick, 2002), which is often thoughts about the self (Dunning, Meyerowitz, & Holzberg, 2002; Kruger, 1999). We over-rely on our

introspections, considering the more objective, behavioral facts as secondary (Pronin & Kugler, 2007). We are biased toward self-enhancement and justify our beliefs and actions—even altruistic action—in terms of self-interest, (Miller, 1999). Even beyond the "better-than-average" effect, our thinking is biased toward the self in comparison to others (Sedikides & Gregg, 2008). This is not mere egoism, for while we tend to think of ourselves as "better than average," we can also think of ourselves as below average in comparison to others (Kruger, 1999). We consider our thoughts as representative of a reality that is shared by others (Gilbert & Malone, 1995; Ross & Ward, 1996). Moreover, we tend to assume too early that our memories, judgments, intuitions, and beliefs are sufficient for the epistemic task at hand (Evans, 2007).

Self-centered thinking is not, in-and-of-itself, IA. Indeed, it is only natural that our own experiences are going to be more readily available to us as evidence. It can only be characterized as arrogance when self-centered thinking is not sufficient to hold a belief in accordance with the evidence, if we count on what we know more than we ought.[11] Normally, Type 1 processing provides the information needed to make the decisions and judgments necessary to successfully navigate life (Evans, 2007; Stanovich, 1999). It does so with great efficiency. The problem comes when Type 1 processes are not sufficient and thinking is not decoupled from Type 1 representations: when it is necessary to move beyond what one immediately knows, believes, and/or remembers to incorporate other information. That is, when Type 2 processes intervene and, in a slower, deliberative, and sequential manner, bring more information to bear until the mind is satisfied that it knows what it needs to know. That information can come from within the self, by accessing memories which broaden the data or evidence at hand, it can come from others who offer

a different perspective, and it can come from the application of rules of analysis that help determine what is sufficient to make a judgment. Type 2 processing, however, is also prone to bias because of the sequential nature of its processing, taking up one mental model at a time until a sufficient one is found (Evans, 2007). Some biases, like those that occur from framing effects, can be the result of a Type 2 processing that has decoupled, but is focused on a singular model from a source outside the self that stimulates all other subsequent thought (serial associative cognition with a focal error, Stanovich, 2009). This might be best characterized as an exhibition of intellectual diffidence (ID)– the opposite of IA because it is too 'other' focused–yet still missing mark of holding a belief with the proper firmness that the belief warrants.[12]

The examples of behaviors and techniques noted above that help to reduce biases share the common aspect of "de-centering" (decoupling) thought to include and consider the "other" through the use of rules of reasoning, a concern for accuracy, a need for accountability, and taking the perspective of others. One common theme in the literature is the role of motivation in de-biasing thought (Chaiken, et al., 1996; Dunning, et al., 1995; Kunda, 1990; Wilson, et al., 2002), pointing to the conclusion correcting biases is an effortful process that requires some kind of motivation to overcome the self-centered tendencies of our cognitive system. In this way, it reflects the neo-Aristotelian (responsibilist) notion of epistemic virtue: that it must be consciously practiced to overcome our more arrogant cognitive tendencies and avoid the possibility of being too diffident to others by giving in too easily or not evaluating the other's position rigorously.

§4.1. Avoiding and mitigating biases with intellectual humility.

We propose that the set of traits and dispositions that best avoid the vices of IA and ID as we have defined them and help mitigate biases (open-mindedness, gathering all available evidence, concern for accuracy and accountability, among others) is best characterized by the epistemic virtue of intellectual humility (IH). We have seen evidence that epistemic characteristics and behaviors that might contribute to IH are both influenced by situations (inducing IH as a state, Armor & Taylor, 2002) and part of an individual's thinking dispositions (which makes IH look more trait-like, Stanovich & West, 1997, Stanovich, 2009). If IH requires the effortful control of Type 2 thinking processes, it may be hard to defend IH as a trait that is held with any stability, making IH appear more state than trait. Never-the-less, there are individual differences in the capacity for effortful control and the characteristics that lend themselves to IH that are more stable and trait-like. Part of the resolution might be found in Stanovich's tri-partite mind (Stanovich, 2009, Evans & Stanovich, 2013), which posits an autonomous mind (Type 1 processes), an algorithmic mind, and a reflective mind (two aspects of Type 2 processes). The algorithmic mind, being highly correlated with general intelligence, is more trait-like in its manifestations, while the reflective mind is defined by state-like thinking dispositions (though these can also show trait-like individual differences in certain conditions [see Stanovich & West, 2007]). Either way, responsibilist virtue epistemology would recognize that IH *can* be developed, enhanced, and trained. On the one hand, those low in the trait could cultivate the habits of IH through practice (while those high in the trait could grow in expertise through the same means). On the other, creating the right pressures and circumstances (accountability, etc.) could induce IH (or attenuate IA and ID).[13] Like the state/trait debate in personality and social psychology,

IH is neither simply state or trait but result of the complex interaction of person and environment that is a part of all personality development (Fleeson, 2004).

From a dual-process point of view, another empirical question worthy of investigation is whether or not "other-centered" thinking could become a part of Type 1 cognition, i.e., whether or not it could become a habit of thought which is automatic, effortless, and intuitive. Some indications in the literature suggest that it can. For example, the short intervention in perspective taking outlined above resulted in implicit and unconscious correction of racial bias (Todd, et al., 2011). Thinking dispositions and algorithmic thinking can be trained and practiced to automaticity (Stanovich, 2009). The literature on moral expertise could also yield clues about the development of habitual, "other-centered" thinking. Using insights from dual process theory, Narvaez (2008) sees both systems at work in the development of moral expertise noting that expertise is trained by immersing novices in environments that build up their intuitions and by giving explicit guidance on how to solve problems in the given domain. The training of moral intuitions begins even in the early experiences of attachment and continues through novice-to-expert instruction. Drawing on expertise training in other fields, Narvaez notes that in such training "perceptions are fine-tuned and developed into chronically accessed constructs, applied automatically; action schemas are honed to high levels of automaticity" (p. 313). Like developing any virtuous habit, intentionality on the part of parents, teachers, and mentors to train habits such as open-mindedness, accurate representation of reality, and the search for the best evidence in a systematic and sustained manner would contribute to the "chronic accessibility" of IH.

§5. Conclusion

While it may seem as though the virtue of intellectual humility might be best explicated in terms neo-Aristotelianism virtue epistemology's character traits, there is no obvious reason why our concepts of intellectual humility cannot be explicated the terms of agent-relaibilism virtue epistemology as well. Presumably, any robust, full account of intellectual virtue will have to account for both cognitive faculty-virtues as well as character trait-virtues; whether one does this along agent-reliabilism lines or neo-Aristotelianism lines could be, to some extent, a matter of emphasis. For a person can hold a belief more strongly (or weakly) than warranted due to biases inherent in our cognitive systems, or due to some lack of character, just as a person can exhibit virtuous knowing via the proper functioning of one's cognitive system or through the exercise of a virtuous character.

Straightforwardly, those of us interested explicating intellectual humility in light of the function (or malfunction) of our cognitive faculties might take special interest in the model of intellectual virtue afforded by agent-reliabilism. Conversely, those of us interested in explicating intellectual humility as something more like a character trait might take special interest in the model of intellectual virtue afforded by neo-Aristotelianism. Either way, it can be defined as holding a belief as firmly as it is warranted, whether such warrant is derived from the proper functioning of our cognitive systems or whether such warrant is brought about by the exercise of a particular way of knowing (a trait).

Within either framework of virtue epistemology, we argue that the epistemic virtue of intellectual humility is mostly found in the conscious exercise of Type 2 thinking and can

come about in a manner consistent with either mode of virtue epistemology: through the proper collaboration of Type 1 and Type 2 processes (agent-reliabilism) or through the conscious practice of applying Type 2 thinking when appropriate (neo-Aristotelians). The literature investigating ways to attenuate cognitive biases suggests at least four habits that would lend themselves to cultivating intellectual humility and lead away from intellectual arrogance: the use of rules of analysis, a search for accuracy, being accountable for one's judgments, and exposure to differing perspectives. Certain traits appear to create a natural disposition toward intellectual humility and other forms of virtuous knowing (i.e., the need for cognition, [Cacioppo & Petty, 1982], active open minded thinking [Stanovich & West, 1997], the need for closure [Kruglanski, 1990], and those traits from factor models of personality that relate to IH such as open-mindedness, conscientiousness, and humility [Ashton & Lee, 2005; McCrae & Costa, 1997]). These traits can be trained, enhanced, and cultivated. Just as habits play a role in cultivating any virtue, habits of the mind can bring about intellectual humility and other epistemic virtues. They serve to curb the vice of intellectual arrogance, even in the face of a cognitive system that can be prone to biased thinking. Perhaps our cognitive system is not in and of itself "vicious," but it may take conscious effort and virtuous habits of the mind not to make it so.

References

Alfano, M. (2012). Expanding The Situationist Challenge To Responsibilist Virtue Epistemology. *Philosophical Quarterly*, *62*(247), 223–249.

Armor, D. A., & Taylor, S. E. (2002). When predictions fail: The dilemma of unrealistic optimism. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 334-347). New York, NY US: Cambridge University Press.

Ashton, M. C., & Lee, K. (2005). Honesty-humility, the big five, and the five-factor model. *Journal of Personality, 73*(5), 1321-1353. doi: 10.1111/j.1467-6494.2005.00351.x

Axtell, G. (forthcoming). Thinking Twice About Virtue and Vice: From Epistemic Situationism to Dual Process Theories. In *Epistemic Situationism*. Oxford University Press.

Baron, J. (1994). *Thinking and deciding* (2nd ed.). New York, NY US: Cambridge University Press.

Birch, S. A. J., & Bernstein, D. M. (2007). What can children tell us about hindsight bias: A fundamental constraint on perspective-taking? *Social Cognition, 25*(1), 98-113. doi: 10.1521/soco.2007.25.1.98

Blank, H., Fischer, V., & Erdfelder, E. (2003). Hindsight bias in political elections. *Memory, 11*(4-5), 491-504. doi: 10.1080/09658210244000513

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116-131. doi: 10.1037/0022-3514.42.1.116

Chaiken, S., Wood, W., & Eagly, A. (1996). Principles of persuasion. In E. T. Higgins & A. W. Kruglanski (Eds.), *Social Psychology: Handbook of basic principles.* New York: The Guilford Press.

Conee, E., & Feldman, R. (1998). The generality problem for reliabilism. *Philosophical Studies: An International Journal for Philosophy in the Analytic Tradition, 89*(1), 1-29.

Dunning, D., Krueger, J. I., & Alicke, M. D. (2005). The self in social perception: Looking back, looking ahead. In M. D. Alicke, D. A. Dunning & J. I. Krueger (Eds.), *The self in social judgment* (pp. 269-280). New York, NY US: Psychology Press.

Dunning, D., Leuenberger, A., & Sherman, D. A. (1995). A new look at motivated inference: Are self-serving theories of success a product of motivational forces? *Journal of Personality and Social Psychology, 69*(1), 58-68. doi: 10.1037/0022-3514.69.1.58

Dunning, D., Meyerowitz, J. A., & Holzberg, A. D. (2002). Ambiguity and self-evaluation: The role of idiosyncratic trait definition in self-serving assessments of ability. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 324-333). New York, NY US: Cambridge University Press.

Epley, N., & Gilovich, T. (2002). Putting adjustment back in the anchoring and adjustment heuristic. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 139-149). New York, NY US: Cambridge University Press.

Epstein, S., Lipson, A., Holstein, C., & Huh, E. (1992). Irrational reactions to negative outcomes: Evidence for two conceptual systems. *Journal of Personality and Social Psychology, 62*(2), 328-339. doi: 10.1037/0022-3514.62.2.328

Evans, J. T. (2007). *Hypothetical thinking: Dual processes in reasoning and judgment.* New York, NY: Psychology Press.

Evans, J. T., & Stanovich, K. E. (2013). Dual-process theories of higher cognition: Advancing the debate. *Perspectives On Psychological Science, 8*(3), 223-241.

Fleeson, W. (2004). Moving personality beyond the person-situation debate: The challenge and the opportunity of within-person variability. *Current Directions in Psychological Science, 13*(2), 83-87. doi: 10.1111/j.0963-7214.2004.00280.x

Gilbert, D. T., & Malone, P. S. (1995). The correspondence bias. *Psychological Bulletin, 117*(1), 21-38. doi: 10.1037/0033-2909.117.1.21

Gilovich, T. (1991). *How we know what isn't so: The fallibility of human reason in everyday life.* New York, NY US: Free Press.

Gilovich, T., Griffin, D., & Kahneman, D. (Eds.). (2002). *Heuristics and biases: The psychology of intuitive judgment.* New York, NY US: Cambridge University Press.

Greco, J. (2003). Virtue and luck, epistemic and otherwise. *Metaphilosophy, 34*(3), 353-366.

Greco, J. (2009). Knowledge and Success from Ability. *Philosophical Studies*, *142*(1), 17–26.

Greco, J. (2010). *Achieving knowledge: A virtue-theoretic account of epistemic normativity.* Cambridge: Cambridge University Press.

Greco, J. (2012). A (Different) Virtue Epistemology. *Philosophy and Phenomenological Research*, *85*(1), 1–26.

Guenther, C. L., & Alicke, M. D. (2010). Deconstructing the better-than-average effect. *Journal of Personality and Social Psychology, 99*(5), 755-770. doi: 10.1037/a0020959

Hergovich, A., Schott, R., & Burger, C. (2010). Biased evaluation of abstracts depending on topic and conclusion: Further evidence of a confirmation bias within scientific psychology. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues, 29*(3), 188-209. doi: 10.1007/s12144-010-9087-5

Jonas, E., Schulz-Hardt, S., Frey, D., & Thelen, N. (2001). Confirmation bias in sequential information search after preliminary decisions: An expansion of dissonance theoretical research on selective exposure to information. *Journal of Personality and Social Psychology, 80*(4), 557-571. doi: 10.1037/0022-3514.80.4.557

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

Kahneman, D., & Frederick, S. (2002). Representativeness revisited: Attribute substitution in intuitive judgment. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 49-81). New York, NY US: Cambridge University Press.

Kruger, J. (1999). Lake Wobegon be gone! The 'below-average effect' and the egocentric nature of comparative ability judgments. *Journal of Personality and Social Psychology, 77*(2), 221-232. doi: 10.1037/0022-3514.77.2.221

Kruglanski, A. W. (1990). Lay epistemic theory in social-cognitive psychology. *Psychological Inquiry, 1*(3), 181-197. doi: 10.1207/s15327965pli0103_1

Kruglanski, A. W., Dechesne, M., Orehek, E., & Pierro, A. (2009). Three decades of lay epistemics: The why, how, and who of knowledge formation. *European Review of Social Psychology, 20*(1), 146-191. doi: 10.1080/10463280902860037

Kruglanski, A. W., & Mayseless, O. (1987). Motivational effects in the social comparison of opinions. *Journal of Personality and Social Psychology, 53*(5), 834-842. doi: 10.1037/0022-3514.53.5.834

Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin, 108*(3), 480-498. doi: 10.1037/0033-2909.108.3.480

McCrae, R. R., & Costa, P. T., Jr. (1997). Personality trait structure as a human universal. *American Psychologist, 52*(5), 509-516. doi: 10.1037/0003-066x.52.5.509

Mercier, H., & Sperber, D. (2011). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences, 34*(2), 57-74. doi: 10.1017/s0140525x10000968

Narvaez, D. (2008). Human flourishing and moral development: Cognitive and neurobiological perspectives of virtue development. In L. P. Nucci & D. Narvaez (Eds.), *Handbook of moral development* (pp. 310-327). New York: Routledge.

Nickerson, R. S. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175-220. doi: 10.1037/1089-2680.2.2.175

Olin, L., & Doris, J. M. (2013). Vicious Minds. *Philosophical Studies*, 1–28. doi: 10.1007/s11098-013-0153-3

Petty, R. E., Wegener, D. T., & White, P. H. (1998). Flexible correction processes in social judgment: Implications for persuasion. *Social Cognition, 16*(1), 93-113. doi: 10.1521/soco.1998.16.1.93

Plantinga, A. (1993). *Warrant: The current debate.* New York, NY: Oxford University Press.

Pritchard, D. (2005). *Epistemic luck*. Oxford, UK: Oxford University Press.

Pronin, E., Berger, J., & Molouki, S. (2007). Alone in a crowd of sheep: Asymmetric perceptions of conformity and their roots in an introspection illusion. *Journal of Personality and Social Psychology, 92*(4), 585-595. doi: 10.1037/0022-3514.92.4.585

Pronin, E., & Kugler, M. B. (2007). Valuing thoughts, ignoring behavior: The introspection illusion as a source of the bias blind spot. *Journal of Experimental Social Psychology, 43*(4), 565-578. doi: 10.1016/j.jesp.2006.05.011

Ross, L., & Ward, A. (1996). Naive realism in everyday life: Implications for social conflict and misunderstanding. In E. S. Reed, E. Turiel & T. Brown (Eds.), *Values and knowledge* (pp. 103-135). Hillsdale, NJ England: Lawrence Erlbaum Associates, Inc.

Sanna, L. J., Schwarz, N., & Stocker, S. L. (2002). When debiasing backfires: Accessible content and accessibility experiences in debiasing hindsight. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*(3), 497-502. doi: 10.1037/0278-7393.28.3.497

Sedikides, C., & Gregg, A. P. (2008). Self-enhancement: Food for thought. *Perspectives on Psychological Science, 3*(2), 102-116. doi: 10.1111/j.1745-6916.2008.00068.x

Sedikides, C., Horton, R. S., & Gregg, A. P. (2007). The why's the limit: Curtailing self-enhancement with explanatory introspection. *Journal of Personality, 75*(4), 783-824. doi: 10.1111/j.1467-6494.2007.00457.x

Sloman, S. A. (2002). Two systems of reasoning. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 379-396). New York, NY US: Cambridge University Press.

Stanovich, K. E. (1999). *Who is rational?: Studies of individual differences in reasoning.* Mahwah, NJ US: Lawrence Erlbaum Associates Publishers.

Stanovich, K. E. (2009) Distinguishing the reflective, algorithmic, and autonomous minds: Is it time for a tri-process theory? In St. B. T. Evans & K. Frankish (Eds.), *In two minds: Dual processes and beyond* (pp. 55-88). New York, NY US: Oxford University Press.

Stanovich, K. E., & West, R. F. (1997). Reasoning independently of prior belief and individual differences in actively open-minded thinking. *Journal of Educational Psychology, 89*(2), 342-357. doi: 10.1037/0022-0663.89.2.342

Stanovich, K. E., & West, R. F. (1998). Individual differences in rational thought. *Journal of Experimental Psychology: General, 127*(2), 161-188. doi: 10.1037/0096-3445.127.2.161

Stanovich, K. E., & West, R. F. (2002). Individual differences in reasoning: Implications for the rationality debate? In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 421-440). New York, NY US: Cambridge University Press.

Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive

ability. *Thinking & Reasoning, 13*(3), 225-247. doi:10.1080/13546780600780796

Todd, A. R., Bodenhausen, G. V., Richeson, J. A., & Galinsky, A. D. (2011). Perspective taking combats automatic expressions of racial bias. *Journal of Personality and Social Psychology, 100*(6), 1027-1042. doi: 10.1037/a0022308

Wegener, D. T., & Petty, R. E. (1997). The Flexible Correction Model: The role of naive theories of bias in bias correction. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (vol. 29, pp. 141-208). London: Academic Press.

Wilson, T. D., Centerbar, D. B., & Brekke, N. (2002). Mental contamination and the debiasing problem. In T. Gilovich, D. Griffin & D. Kahneman (Eds.), *Heuristics and biases: The psychology of intuitive judgment* (pp. 185-200). New York, NY US: Cambridge University Press.

Zagzebski, L. (1996). *Virtues of the mind.* Cambridge: Cambridge University Press.

---

[1] Some philosophers worry that psychological theory threatens the viability of virtue epistemology (Olin & Doris 2013; Alfano 2012). While not aimed at engaging with that literature, this paper does elucidate a surprising harmony between the philosophical and psychological research—a harmony that can, if anything, help dissolve such worries.

[2] To be sure, one need not be committed to virtue epistemology per se to account for intellectual virtues – someone can account for intellectual virtues without any special loyalty to virtue epistemology – nevertheless, virtue epistemology, naturally enough, offers the most robust and flourishing accounts of intellectual virtue in the philosophical literature.

[3] As Earl Conee and Richard Feldman noted in "The Generality Problem for Reliabilism" (1998), "A fully articulated [process] reliabilist theory must identify with sufficient clarity the nature of the processes it invokes. In doing so, the theory confronts what has come to be known as 'the generality problem'" (1998, p. 1).

[4] Also quoted in Pritchard (2005), p. 188.

[5] See Pritchard (2005), p. 188; Greco (2003), pp. 356-357.

[6] See chapter 7 of Evans (2007) for a comparative analysis.

[7] Recently the language of "systems" has been abandoned by some cognitive scientists in favor of using Type 1 and Type 2 to distinguish these cognitive processes because they do no represent a single system but the many cognitive systems and neural networks that support each type of thinking (Evans & Stanovich, 2013).

[8] Included in type 1 processing are implicit learning and conditioning, decision-making rules and principles that are so well practiced as to be automatic, and the regulation of behavior by emotions. Stanovich (2009) has given the name TASS (The Autonomous Set of Systems) to these processes because they "respond automatically to triggered stimuli [and are not] under the control of the analytic processing system (System 2)" (p. 57).

[9] And this means that according to the credit theory of knowledge espoused by some reliabilist virtue epistemologists (see Greco 2009; Greco 2010; Greco 2012), Type 1 cognition can produce credit-worthy beliefs—beliefs we would deem knowledge—so long as it is employed in a context where it is generally adequate to its task. Also see Axtell forthcoming.

[10] As measured by SAT scores and a test of verbal ability, though cognitive ability was also a unique and independent predictor of argument evaluation performance.

[11] Since much of Type 1 processes are unconscious, we are not accusing the agent of being willfully arrogant. Instead arrogance means to preference the self as a source of information, when what the self knows, by itself, does not provide enough for believing in accordance with the facts, whether the agent is conscious of this preference or not.

[12] We have defined elsewhere that intellectual humility as a virtuous mean that can be defined as "holding a belief with the firmness warranted," which avoids the vice of intellectual arrogance (holding your belief too firmly when it is not warranted) on the one hand, and intellectual diffidence (holding a belief too loosely or giving in to another's belief too soon) on the other (Authors, 2012).

[13] Insofar as research into heuristics and biases helps us understand IH, it is useful to think of IH as the absence or the opposite of IA. As we state above in the introduction, we hold the view that IH is not simply

the opposite, or absence of something (like IA) but that a robust definition should include positive attributes.

This paper was made possible by a generous grant from the John Templeton Foundation to study the Science of Intellectual Humility.